

# Språk og skrift som er brukt i SOS3003

Erling Berge

Spring 2010

© Erling Berge 2010

1

Ei typisk setning i regresjonsspråket:

- $Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$  ,  $i=1, \dots, n$
- Det vi må lære først er rett å slett å lese ei setning som ser ut som den ovanfor.

Spring 2010

© Erling Berge 2010

2

## Her er tolkinga av dei ulike symbola:

- $Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$
- $i=1, \dots, n$
- $n = \#$  case i populasjonen
- $\varepsilon_i$  er feilleddet til case nr  $i$
- $x_{1i}$  er verdien som case nr  $i$  har på variabelen  $x_1$
- $Y_i$  er verdien som case nr  $i$  har på variabelen  $Y$
- $\beta_0$  og  $\beta_1$  er verdiane til regresjonsparametranne  $\beta_0$  og  $\beta_1$  i populasjonen det  $Y$  og  $x_1$  er målt
- $+$  og  $=$  er symbol som er henta frå og tyder det same som i vanleg rekning med tal

Spring 2010

© Erling Berge 2010

3

## Skrift og skriftsymbol

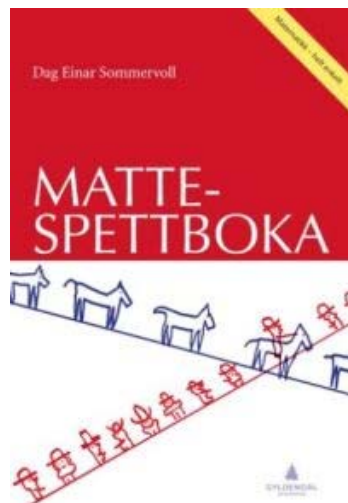
- I SOS3003 møter vi mange nye omgrep
- Somme av omgrepa har ein laga særlege symbol for, mest for å sleppe å skrive lange setningar med definisjonar
- Vi bør sjå på dette som å lære eit nytt språk. Skal vi lære eit nytt ord er det berre ein måte å gjere det på: hugse kva det tyder!
- Men først: ein del ting kan vere nyttig å repetere:

Spring 2010

© Erling Berge 2010

4

## Mattespettboka av Dag Einar Sommervoll (Gyldendal, 2009)



Spring 2010

© Erling Berge 2010

5

### Reknereglar

- Brøkrekning
- Bruk av parentesar
- Likningar med 1 og 2 ukjende
- Relative storleikar
- Funksjonar

## Vanleg rekning med symbol

- Vanleg rekning med symbol er kalla matematikk
- Matematikk er **ikkje** vanskeleg
- Det er eit språk for logikken.
- Det er lett å lære og å lese
- Det kan vere litt vanskelegare å forstå det ein les og
- Enno litt vanskelegare å skrive det sjølv
- Den "dialekten" vi brukar i SOS3003, regresjonsanalyse, er enkel å lese og forstå
- Den er i mange samanhengar omtala som algebra

Spring 2010

© Erling Berge 2010

6

## Litt om kva vi treng

- **Algebra:** rekning med "bokstavar", dvs. vilkårlege tal, inklusiv **potensar og logaritmar**
- **Funksjonar:** prosedyrar for å binde saman verdien av ein type tal (t.d.  $\langle x = \text{År med utdanning} \rangle$  med verdien av ein annan type tal t.d.  $\langle y = \text{Inntekt i kroner} \rangle$ ; Funksjonen  $f$  er da prosedyren som bind saman talet  $x$  med talet  $y$ . Vi skriv  $y=f(x)$
- **Likningar:** påstandar om at to matematiske uttrykk logisk sett er identiske

## Funksjon eller likning?

- Legg merke til skilnaden mellom funksjon og likning. Dei kan sjå like ut på papiret, men likskapsteiknet har to ulike meiningar
- I **funksjonen**  $y=f(x)$  tyder "=" at vi skal gjere  $y$  lik  $f(x)$ . Skal vi klare det må  $f$  vere kjent
- I **likninga**  $y=f(x)$  tyder "=" at  $y$  er identisk lik  $f(x)$ . Dette kan nyttast til å fastleggje funksjonen  $f$  eintydig

## Om konvensjonar for symbolbruk

- Konvensjonell symbolbruk varierer frå fag til fag og frå miljø til miljø
- Innan vår type matematikk (multivariat regresjon) vil ein vanlegvis skilje symbolbruken for ulike typar storleikar:
  - "bokstaval", parametrar, konstantar: a, b, c, d, e, p, q, r, s, t, ... , og tilsvarande i greske bokstavar:  $\alpha$ ,  $\beta$ ,  $\chi$ ,  $\delta$ ,  $\varepsilon$ ,  $\pi$ ,  $\theta$ ,  $\rho$ ,  $\sigma$ ,  $\tau$ , ...
  - variablar vil som regel nytte symbola x, y, z, ...
  - indeksar for variablar nyttar gjerne i, j, k, m, n
  - funksjonar vil ofte bli gitt symbola f(), g(), ...
- Det løner seg som regel å vere eksplisitt i definisjonen av kva symbola tyder

## Algebra (1)

- Latinske bokstavar a,b,c,... står for eit vanleg vilkårleg tal (inklusive negative tal) (Hugs:  $a = +a$ )
- Reknereglar og symbola deira:
  - Addisjon: symbol + t.d.  $a+b = b+a$
  - Subtraksjon symbol - t.d.  $a-b = -b+a$
  - Multiplikasjon symbol \* t.d.  $a*b = b*a$
  - Divisjon symbol / t.d.  $a/b = a*(1/b)$
- Algebraiske uttrykk vert sett saman og ordna ved hjelp av reknereglane og parentesar

## Algebra (2)

- Rekning med parentesar:
- Det som står inne i parentesen kan handsamast som eitt tal.

$$\text{Sett } b+c = t, \text{ da er } a*(b+c) = a*t$$

- Alle element inne i ein parentes må handsamast likt

$$a*(b+c) = a*b + a*c$$

Eks:

$$\begin{aligned}(a+b)*(c+d) &= a*(c+d) + b*(c+d) \\ &= a*c+a*d+b*c+b*d\end{aligned}$$

Spring 2010

© Erling Berge 2010

11

## Algebra (3) Potensar

- Potensar, eksponentar, og logaritmer er i prinsippet same fenomenet. Dei er alle ei form for multiplikasjon.

- Potensar:

$$a*a = a^2$$

$$a^2*a = a^3$$

osv.

$$a*a*a*a* \dots *a \text{ (n gonger)} = a^n$$

- $a^0 = 1$  ; pr definisjon for alle  $a \neq 0$
- $a^{-n} = 1/a^n$  ; pr definisjon
- $a^{1/n} = \sqrt[n]{a}$  ; [ n-te rota av a ]

Spring 2010

© Erling Berge 2010

12

## Reknereglar for potensar

- $a^m \cdot a^n = a^{m+n}$
- $(a \cdot b)^n = a^n \cdot b^n$
- $(a^m)^n = a^{m \cdot n}$
- $a^m / a^n = a^{m-n}$
- $(a/b)^n = a^n / b^n$
  
- I somme samanhengar vert potensen kalla eksponent, og potensering heiter da eksponentiering. Ein spesielle type potens vert kalla logaritmar.

Spring 2010

© Erling Berge 2010

13

## Logaritmar

- Logaritmar er potensar definert i høve til eit grunntal
  - Vanlegvis er grunntalet "10" eller "e"
  - Logaritmar med grunntalet "10" vert kalla Briggske logaritmar: t.d.  $\log(a)$  er logaritmen til a med grunntal 10
  - Grunntalet "e" gir naturlege logaritmar: t.d.  $\ln(a)$
  - Men logaritmar kan i prinsippet definerast vilkårleg i høve til eit grunntal g: t.d.  $\log_g(a)$  er logariment til a med grunntal g
- Logaritmen til talet "a" er definert som eksponenten (eller potensen) grunntalet må opphøggjast i for å få talet "a".  
Dvs.
  - $a = 10^{\log(a)}$
  - $a = e^{\ln(a)} = \exp\{\ln(a)\}$
  - $a = g^{\log_g(a)}$
- Vi skal om ikkje anna vert sagt nytte naturlege logaritmar

Spring 2010

© Erling Berge 2010

14

## Reknereglar for logaritmar

- $\ln(a \cdot b) = \ln(a) + \ln(b)$
- $\ln(a/b) = \ln(a) - \ln(b)$
- $\ln(a^n) = n \cdot \ln(a)$
- $\ln(a^{-n}) = (-n) \cdot \ln(a)$
- $\ln(a^{1/n}) = (1/n) \cdot \ln(a)$

## Funksjonar (1)

- Ein funksjon er ei samling prosedyrar som fortel kva for utfall (eitt eller fleire tal) som høyrer saman med ein gitt input (eitt eller fleire tal). Prosedyresamlinga kan kallast  $f(x)$ ,  $g(x)$ , der " $x$ " står for eit eller anna algebraisk uttrykk
- T.d.: til ein gitt  $x$  som input svarar ein bestemt  $y$  (eller  $f(x)$ ) som utfall
- Med berre ein input-storleik og ein utfallstorleik vil samanhengen kunne framstillast grafisk. Vi talar om grafen til funksjonen



## Merknad om skrivemåte

- Indeksering nyttar vi for å skilje frå kvarandre ulike storleikar av same type
- Indeksar vert gjerne plassert som fotskrift (subskript) etter typen storleik
- T.d.  $f_1$  eller  $f_2$  for to ulike funksjonar
- Men dei kan og plasserast framanfor både som fotskrift og toppskrift. Dette er vanleg når det trengst fleire ulike indeksar
- T.d.  ${}_1g_1$  eller  ${}^1g_2$
- Indeksar kan vere bokstavgang som t.d. i, j, k, ...

Spring 2010

© Erling Berge 2010

17

## Funksjonar (2)

- Eksempel:
- Funksjonen  $f_1(x)$ : til verdien  $x$  svarar utfallet  $f_1(x)$   
 $x$  : 1 : 2 : 3 : 4 : 5 : 6 : 7 : ...  
 $f_1(x)$  : 2 : 4 : 6 : 8 : 10 : 12 : 14 : ...
- Funksjonen  $f_2(x)$  til verdien  $x$  svarar utfallet  $f_2(x)$   
 $x$  : 1 : 2 : 3 : 4 : 5 : 6 : 7 : ...  
 $f_2(x)$  : 1 : 4 : 9 : 16 : 25 : 36 : 49 : ...
- Lag grafar for funksjonane  $f_1$  og  $f_2$

Spring 2010

© Erling Berge 2010

18

## Matematiske operatorar

- For å lette skrivearbeidet med addisjon og multiplikasjon tar vi i bruk operatorane (dette kan og sjåast spesifikt definerte funksjonar):
- Summering:  $\Sigma$ 
  - $a_1 + a_2 + a_3 + a_4 + \dots + a_n = \Sigma_i a_i$
  - ( $= \Sigma a_i$  dersom ikkje anna er sagt)
- Multiplikasjon:  $\Pi$ 
  - $a_1 * a_2 * a_3 * a_4 * \dots * a_n = \Pi_i a_i$
  - ( $= \Pi a_i$  dersom ikkje anna er sagt)
- Eksponentiering:  $\exp\{x\}$ 
  - $Y = \exp\{x\}$  tyder at talet "e" skal opphøgjast i potensen x (eller eksponenten x)
  - Talet  $e=2.71828\dots$  er eit viktig tal (slik som  $\pi = 3.14\dots$ )

## Likningar (1)

- Likningar får vi når vi veit eller kan argumentere for at to algebraiske uttrykk, eller to funksjonar, eller ein funksjon og eit algebraisk uttrykk er identisk like
  - t.d.  $y = f(x)$
  - eller  $y = a + b*x$  (likninga for ei linje)
  - eller  $0 = a + b*x + c*x^2$  (andregradslikninga)
  - eller  $y_i = b_0 + b_1*x_{i1} + b_2*x_{i2} + e_i$   
(regresjonslikning med 2 x-variablar)

## Likningar (2)

- Likningar som inneheld ukjente storleikar kan vere til hjelp i å finne kva dei ukjente storleikane er, anten eksakt eller tilnærma
- Løysinga kan vere algebraisk  

$$0 = a + b \cdot x \Rightarrow x = -a/b$$
- Løysinga kan vere grafisk  
 – t.d. finn x-verdiane som svarar til  $y=0$  i grafen til  $y = a + b \cdot x + c \cdot x^2$
- Løysinga kan vere numerisk

Spring 2010

© Erling Berge 2010

21

## Likningar (3)

Rekning i likningar

- I likninga  $y=f(x)$  kan vi
  - Addere eller subtrahere same tal på begge sider av likskapsteiknet
    - Dvs.  $y+a=f(x)+a$  eller  $y-a=f(x)-a$
  - Multiplisere eller dividere med same tal på begge sider av likskapsteiknet
    - Dvs.  $Y \cdot a=f(x) \cdot a$  eller  $y/a=f(x)/a$
- Dette gjer vi for å finne algebraiske uttrykk der vi har ein avhengig ukjent på venstre sida og enten berre kjente storleikar eller uavhengige variablar og kjente storleikar på høgre sida. Dette vil som regel lette drøftinga av samanhengen mellom uavhengige og avhengig variabel

Spring 2010

© Erling Berge 2010

22

## Regresjonslikningar

- Bivariat  $y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$ 
  - Tolk symbolbruken
- Multippel regresjon med n uavhengige var
  - $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_n x_{ni} + \varepsilon_i$
  - $y_i = \beta_0 + \sum_k (\beta_k x_{ki}) + \varepsilon_i$
- Logistisk regresjon er basert på funksjonen
  - $y_i = \alpha / (1 + \gamma \cdot \exp\{-\beta x_i\}) + \varepsilon_i$  med  $\alpha=1$  og  $\gamma=1$  for å kunne tolke y som estimat av eit sannsyn (oftast skrive som ein p med  $\hat{\phantom{p}}$  over)

Spring 2010

© Erling Berge 2010

23

## Logaritmar og logistisk regresjon

- Logistisk regresjon med n uavhengige variablar er basert på funksjonen
 
$$y_i = [1 / (1 + \exp\{-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_n x_{ni})\})] + \varepsilon_i$$
- Logiten er definert som
 
$$L_i = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_n x_{ni})$$
 slik at
 
$$y_i = 1 / (1 + \exp\{-L_i\}) + \varepsilon_i$$
- Vi ser at logiten er ein sum av eksponentar for grunntalet e
- Uttrykket  $O_i = \exp\{L_i\}$  vert kalla Oddsen

Spring 2010

© Erling Berge 2010

24

## Meir om konvensjonar for symbolbruk

- I testing av multivariate modellar vil ein vanlegvis nytte greske bokstavar for storleikar som karakteriserer heile **populasjonen**:  $\alpha, \beta, \chi, \delta, \varepsilon, \phi, \gamma, \dots$
- Tilsvarande storleikar i **utvalet** får latinske bokstavar: a, b, c, d, e, f, g, ..
- Dette er nytting når vi skal drøfte i kva grad det vi observerer i utvalet kan seie noko om tilhøva i populasjonen

Spring 2010

© Erling Berge 2010

25

## Variablar

Matematikarar talar om

- Kontinuerlege (gradvis og jamn vekst/minke i verdiar)
- Diskrete (kan alltid omformast til verdiane 1,2,... )
- Binære (tar verdiane 0 eller 1)

Samfunnsvitarar talar om

- Variablar med måleskala (kroner, kilo, år, etc), dvs. **intervall- eller høvestalskala**
- Variablar med rangeringar, dvs **ordinalskala**
- Klassifikasjonar (sosialklasse, yrke, etc), inklusive dikotomiar (kjønn, regjeringsmakt, etc), dvs. **nominalskala**

Spring 2010

© Erling Berge 2010

26